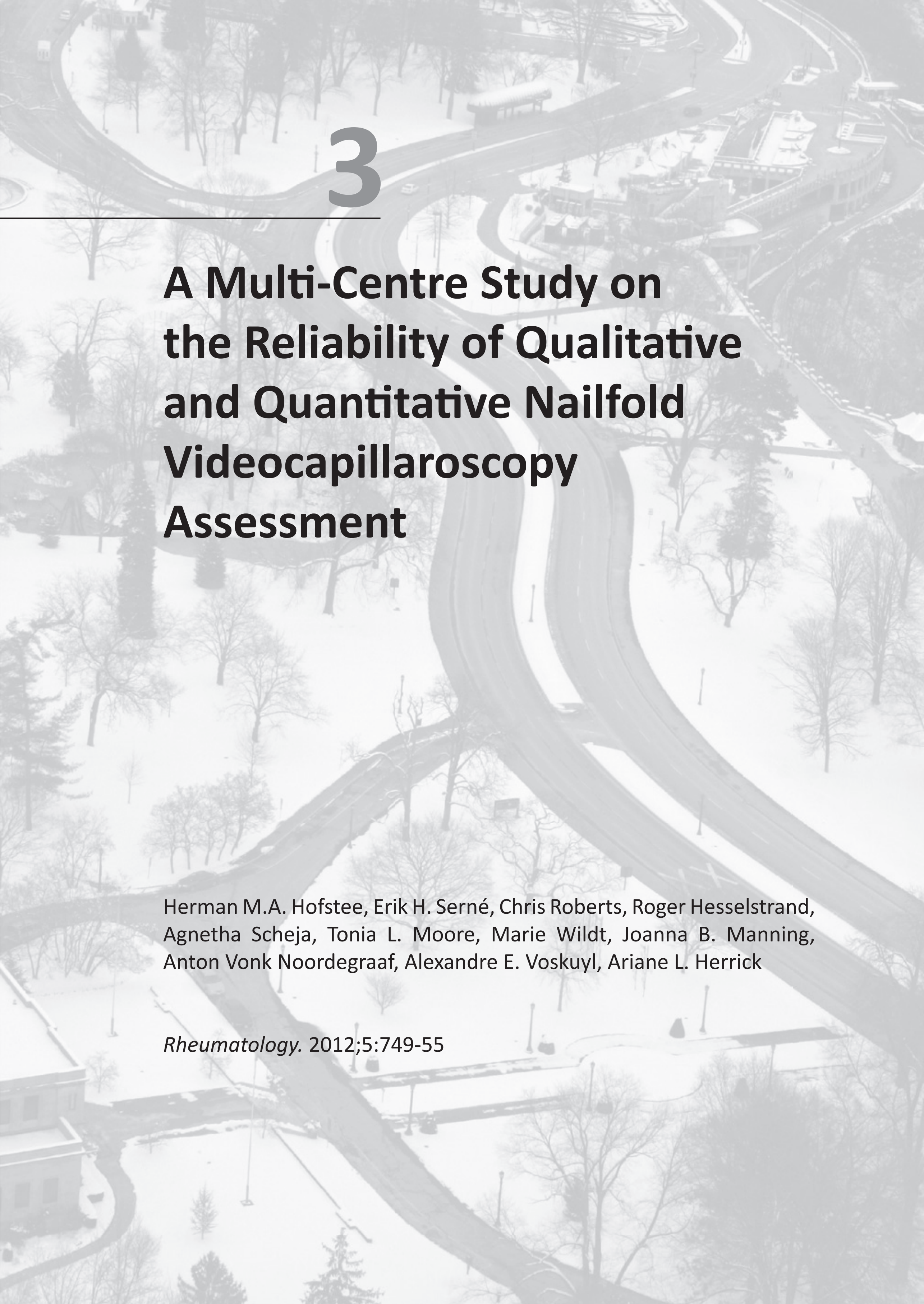
Structural and Functional Microcirculatory Aspects of Patients with (Suspected) Connective Tissue Disease and Pulmonary Arterial Hypertension

Hofstee, H.M.A.

2013



3

A Multi-Centre Study on the Reliability of Qualitative and Quantitative Nailfold Videocapillaroscopy Assessment

Herman M.A. Hofstee, Erik H. Serné, Chris Roberts, Roger Hesselstrand, Agnetha Scheja, Tonia L. Moore, Marie Wildt, Joanna B. Manning, Anton Vonk Noordegraaf, Alexandre E. Voskuyl, Ariane L. Herrick

Rheumatology. 2012;5:749-55

ABSTRACT

Objective: To investigate the inter- and intra-observer reliability of both qualitative and quantitative parameters used in the assessment of nailfold capillaroscopy images.

Methods: Fifty mosaic nailfold images of healthy controls (n=10), patients with primary Raynaud's phenomenon (n=10), and systemic sclerosis (SSc) (n=30) were assessed in random order by two blinded observers on two occasions at centres in Sweden, United Kingdom, and the Netherlands. Each image was therefore scored by 6 observers twice.

Results: Inter- and intra-observer reliability of quantitative parameters showed substantial to almost perfect agreement (inter-and intra-observer weighted kappa's for the number of widened capillaries 0.75 and 0.87 and giant capillaries 0.84 and 0.92, intraclass correlation coefficients (ICCs) for capillary density 0.87 and 0.92, and total loop width 0.94 and 0.98). Qualitative parameters including architecture, avascularity, haemorrhage, crossed, ramified, and bushy capillaries showed moderate to substantial inter-observer reproducibility (weighted kappa ranging from 0.47 to 0.73), and substantial intra-observer repeatability (weighted kappa ranging from 0.71 to 0.80), whereas the scoring of tortuous and bizarre capillaries showed poor inter-observer and substantial intra-observer agreement (inter-observer weighted kappa's 0.39 and 0.21, and intra-observer weighted kappa's 0.68 and 0.76 respectively).

Conclusions: All quantitative and certain qualitative parameters are highly reliable in terms of inter- and intra-observer agreement. A combination of parameters with the highest reliability should be incorporated into future capillaroscopic scoring systems in studies of prediction and monitoring of SSc-spectrum disorders.

INTRODUCTION

Nailfold capillaroscopy is a non-invasive tool that is increasingly used in the evaluation of patients with Raynaud's phenomenon (RP).[1] Patients with developing or established systemic sclerosis (SSc)-spectrum disorders may show characteristic changes in the nailfold capillary architecture and loop dimensions at the time of, or even years before a connective tissue disease (CTD) diagnosis is made.[2-4] These changes consist of enlargement of capillary loops (widened and giant capillaries), drop-out of capillaries (avascularity) and changes in the capillary morphology and architecture.[5-7] Although nailfold capillaroscopy is easy to perform, the assessment of nailfold changes is less straightforward.[8,9] Traditionally, only qualitative and semi-quantitative assessments could be made, but nowadays with the advent of sophisticated microscopes and computerised techniques, it is possible to combine the high magnification of nailfold videocapillaroscopy with the ability to view the whole nailfold (as a mosaic) and make quantitative measurements.[5] However, data on inter-observer reproducibility and intra-observer repeatability concerning nailfold capillary assessments are scarce and confined to a few observers and centres. Knowledge of the reliability of nailfold measurements is essential for the development of simple and reliable nailfold classifications that can be used in clinical practice, and may help to stimulate research in the field of (complications of) CTD, and response to therapy. The aim of the present study was to assess inter- and intra-observer reliability of both qualitative and quantitative parameters between and within observers and centres in Sweden, the United Kingdom, and the Netherlands. In order to study a wide spectrum of nailfold normality and abnormality, healthy controls, patients with RP and SSc were included. However, it was not an aim of the study to examine differences between subject groups.

METHODS

Nailfold images

Capillaroscopy was performed, and computer-based nailfold panorama mosaic images acquired, as previously described.[5] Based on power calculations (see statistical analysis), a random selection of 50 coded images with a x300 magnification of digit 4 (ring finger) of the non-dominant hand from patients with primary RP (n=10), SSc (n=30) and healthy controls (n=10), taken from two large image databases (UK and the Netherlands), were assessed in random order by two observers in each centre twice with a minimal time interval of two weeks between the first and second assessment. Patients with primary RP had no clinical or serological features of an underlying CTD, patients with SSc were all diagnosed by an experienced rheumatologist, and all controls were healthy without RP. Each nailfold was therefore assessed by 6 observers twice using the same analysis software. All ratings were conducted independently. Participating centres were large teaching hospitals in Sweden, the United Kingdom, and the Netherlands. Ten illustrative nailfold images, not included in the study, were used as training files to aid agreement. The local ethics committee and/or Research and Development Directorate in the Netherlands and the UK gave permission for the exchange of images.

Assessment of nailfold images by each observer

Qualitative assessment. Nailfold changes were semi-quantitatively scored for architecture (degree of distortion of the normal regular capillary pattern), avascularity, haemorrhage, and capillary morphology (tortuous, crossed, ramified, bushy, and bizarre) (figure 1).

Architecture was scored as ‘normal’ if there was a regular or slightly disturbed nailfold pattern, and ‘moderate’ or ‘severe’ if there was a moderately or severely disturbed nailfold pattern respectively. Avascularity was defined as loss of two or more consecutive capillaries and scored as ‘none’, ‘moderate’ (loss of 2-4 consecutive capillaries), and ‘severe’ (loss of >4 consecutive capillaries, or > 2 areas with loss of ≥2 capillaries). Haemorrhage was scored as none, moderate (>2 punctate haemorrhages, not confluent), and severe (confluent area). Capillary morphology was assessed according to the taxonomy proposed by Jones et al.[9], and scored by counting the number of morphological abnormal capillaries (‘0’, ‘1’, ‘2’, or ‘≥3’). A capillary was scored as ‘tortuous’ if the capillary limb curled, but did not cross; the capillary limb of a ‘crossed’ capillary crossed at least once. If the capillary limb branched (like a tree without leafs), the capillary was scored as ‘ramified’ (synonymous with ‘arborised’); ‘bushy’ capillaries showed small buds instead of branches. ‘Bizarre’ capillaries showed large variations of very abnormal branching. To aid the identification, examples of morphological abnormalities (figure 1) were printed on a scoring form. Observers were also asked to rate image quality (as ‘good’, moderate’ or ‘poor’).

Quantitative assessment. Mean apical, arterial, venous and total loop width were based on the measurements of the 5 capillaries with the largest total loop width, as independently identified by each observer. In case of irregular dilated capillary loops, maximal width of each dimension was measured. Individual loop measurements were used to score the number of widened- and giant capillaries (‘0’, ‘1’, ‘2’, or ‘≥3’). A widened capillary was defined as having a total loop width of >90 µm but <150 µm.[10] A giant capillary was defined as having a total loop width of ≥150 µm.[10] Capillary density (number of loops per mm) was calculated by computer from the manually marked loops in the distal row. In some cases (especially when capillary density is high) it can be difficult to define which capillaries belong to the distal row. Therefore, each observer assessed capillary density in two ways, firstly by directly observing the loops and marking the loops that were considered to be distal loops by the observer ('direct observation method'), and secondly by the '90^o method' where a capillary loop was considered to be a distal loop if the apex of the capillary made an angle of $\geq 90^o$ with the apex of its adjacent capillaries (figure 2).

Statistical analysis

Sample size. Practical consideration determined the number of observers that could be involved in the study. With 6 observers (2 from each centre) and a minimum of 33 images the study would have 80% power to detect that an intra-class coefficient of 0.5 was significantly greater than a value of 0.3.[11] We are not aware of methods of sample size estimation for multi-observer weighted kappa statistic, so the planned sample size of images was increased to 50 as estimates of weighted kappa will tend to be less precise than those for the intra-class correlation coefficient.

Reliability analysis. For ordered categorical scores, inter-observer reproducibility and intra-observer repeatability were assessed using a weighted kappa coefficient with quadratic weights. Reproducibility and repeatability of continuous scores were assessed using intra-class correlation (ICC). All confidence intervals were determined using a non-parametric bootstrap re-sampling by image. Statistical analyses were carried out using STATA Release 11 (StataCorp LP Texas, USA). The estimate of intra-observer repeatability has been averaged across the observers so as to utilise all the available data. The interpretation of measures of reliability depends on the purpose and consequence of the measure.[12] Landis and Koch suggested the following informal, and somewhat arbitrary banding scheme for the interpretation of measures of reliability for group level decisions or research purposes: <0 no, 0.00 – 0.40 poor, 0.41 - 0.60 moderate, 0.61 - 0.80 substantial, and 0.81 – 1.00 almost perfect agreement.[13]

RESULTS

Table 1 shows the distribution of the categorical scores and summary statistics for the quantitative measures. Table 2 summarises the inter- and intra-observer reliability.

Qualitative assessment. The first assessment showed substantial inter-observer reproducibility on architecture and avascularity, whereas moderate agreement on haemorrhage was present. At the second assessment, inter-observer agreement on haemorrhage was substantial, and that of architecture and avascularity were moderate, although differences in absolute values were small. Capillary morphology showed substantial inter-observer agreement on bushy capillaries, moderate agreement on crossed and ramified capillaries, and poor agreement on bizarre and tortuous capillaries at the first assessment. The scoring of capillary morphology was similar at the second assessment, although inter-observer agreement fell to moderate for bushy capillaries, and agreement between observers on ramified capillaries rose to substantial. The intra-observer repeatability on the other hand, showed overall substantial agreement on all qualitative parameters, with smaller confidence intervals and almost perfect agreement on bushy capillaries. It should be noted that the confidence intervals for most qualitative assessments are wide, which may be due to majority of assessments being in category 'None' for all assessments, except 'Crossed'.

Quantitative assessment. Capillary density and loop dimensions could not be measured in some cases due to poor image quality . The scoring of the quantitatively assessed widened-and giant capillaries showed substantial and almost perfect inter-observer agreement respectively. Measuring capillary density with the ‘90^o method’ was superior to the ‘direct observation method’, showing almost perfect inter-observer agreement with smaller confidence intervals. Almost perfect inter-observer reproducibility was present on the quantitatively assessed loop dimensions (apex, arterial, venous, and total width) on both first and second assessments. All quantitatively assessed parameters showed almost perfect intra-observer repeatability. Intra-observer agreement on capillary density did not differ between the ‘90^o method’ and the ‘direct observation method’.

DISCUSSION

This international multicentre study on the reliability of nailfold assessment shows that the quantitative scoring of capillary loop dimensions, widened- and giant capillaries, and capillary density is highly reliable, showing substantial to almost perfect inter- and intra-observer agreement. Of the qualitative parameters, architecture, avascularity, haemorrhage, crossed, ramified, and bushy capillaries show moderate to substantial reliability, whereas tortuous and bizarre capillaries show poor reliability. We would therefore have reservations including the descriptors ‘tortuous’ and ‘bizarre’ capillaries in future studies. The study was possible due to technical advances in nailfold videocapillaroscopy over recent years, allowing electronic exchange of images and the unique opportunity to both qualitatively and quantitatively assess the mosaic nailfold images of healthy controls and patients with primary RP and SSc across three different centres, with six ‘blinded’ observers assessing each image on two occasions. The main focus of the study was to investigate reliability of image analysis. Variability in image acquisition was not assessed , being out with the remit of the study, but was minimised by using software that produced digitised mosaic images. The inclusion of healthy controls, patients with primary RP and SSc ensured that the findings of our study are representative of a wide spectrum of nailfold capillary normality and abnormality. However, a limitation of the study was the relatively small sample size, which combined with the fact that most qualitative parameters fell in one ordered category, resulted in quite wide confidence intervals round the point estimation of most qualitative parameters.

Although nailfold capillaroscopy is a popular tool (especially in recent years with an increasing number of related publications), data on inter- and intra-observer reliability concerning nailfold capillary parameters are scarce. Good reliability is a necessary prerequisite for a test to be a valid clinical or research tool. For the practising rheumatologist (who is unlikely to perform quantitative assessments), qualitative capillaroscopic parameters must be reliable in terms of reproducibility and repeatability to allow different clinicians to agree on what is normal and what is abnormal when predicting which patients are likely to have an underlying CTD. For the researcher, quantitative parameters must be reliable to allow further exploration of capillaroscopy as an outcome measure in longitudinal studies of SSc-related microvascular disease, including studies of treatment response.

Our previous studies, being single centre with 2 observers only, including estimates of reliability focused on the ability to quantify nailfold changes (capillary density and loop dimensions)[5,14], and showed improvement of inter-rater reliability if the same 5 capillaries were measured.[5] It was against this background, and the fact that abnormalities in (developing) CTD are characterised by widening of capillary loops, that we choose to measure capillary loop dimensions of five capillaries with the largest total width. However, measurement of capillary dimensions alone will lose important information, because this cannot take into account some of the characteristic abnormalities including haemorrhages, and only indirectly reflects others such as ramification. Thus any assessment of nailfold capillaries ideally needs to incorporate qualitative as well as quantitative parameters, the approach which we adopted in this study.

A recent single-centre study (2 observers)[10] assessed the inter- and intra-observer reliability of both quantitative and qualitative measures in 217 patients with RP with or without underlying connective tissue disease. Capillary disorganisation, avascularity, capillary density and haemorrhages showed substantial inter-observer agreement, but intra-observer agreement concerning inter-capillary distance (a parameter which we did not assess, but which is likely to be inversely related to capillary density) was poor, raising questions about the quantitative aspects of the assessment, which were described in categorical scales rather than as continuous variables as in our study. A recent two-centre study[15] reported the inter- and intra-observer variability between two observers of qualitative nailfold patterns and four semi-quantitative 'hallmark' parameters (capillary loss/density, giant capillaries, haemorrhages, and ramifications) in a group of SSc patients, each hallmark parameter assessed on a categorical (0-3) scale. In this study by Smith et al, the ICCs of these 4 parameters were comparable to those in our study, except for haemorrhage (inter-observer ICC 0.90 versus 0.56 in our study). [15]

In recent years, several studies have stressed the importance of nailfold capillaroscopy in identifying (developing) CTD in patients with RP.[2-4,16] Recent criteria (yet to be validated) for the early diagnosis of SSc, proposed by the European League Against Rheumatism (EULAR) Scleroderma Trial and Research Group[17], highlight the pivotal role of capillaroscopy. It is likely that nailfold capillaroscopy will be increasingly used in clinical practice. To be useful, however, both technical issues (magnification used, and the ability to make quantitative measurements) as well as the question of which parameters to score, have to be considered. It is likely that future scoring systems of nailfold images will incorporate both qualitative and quantitative measures, and will include most if not all of the fingers in each subject because capillary morphology can vary between nailfolds within any one individual. Our study, by examining reliability of both qualitative and quantitative parameters across different centres, gives pointers as to which parameters are likely to be preferable in terms of reliability (these are likely to include widened- and giant capillaries, architecture, and all quantitative parameters) and which should be excluded. The method of measuring capillary density must be carefully defined. Further multicentre studies are needed to further develop and improve scoring systems that can differentiate between RP patients with and without (developing) CTD by examining all digits in a simple and reliable way, and to evaluate capillaroscopy as an outcome measure in longitudinal studies.

Table 1. Distribution of the categorical scores and summary statistics for the quantitative measures

	First assessment					Second assessment				
	Good	**Moderate**	**Poor**		**N**	**Good**	**Moderate**	**Poor**		**N**
Image quality (%)	209 (70)	70 (23)	21 (7)		300	223 (74)	57 (19)	20 (7)		300
	Normal	**Moderate**	**Severe**			**Normal**	**Moderate**	**Severe**		
Architecture (%)	239 (80)	47 (16)	14 (5)		300	232 (77)	52 (17)	16 (5)		300
	None	**Moderate**	**Severe**			**None**	**Moderate**	**Severe**		
Avascularity (%)	238 (79)	54 (18)	8 (3)		300	243 (81)	52 (17)	5 (2)		300
Haemorrhage (%)	264 (88)	28 (9)	8 (3)		300	261 (87)	29 (10)	10 (3)		300
No. of capillaries	**None**	**1**	**2**	**3+**		**None**	**1**	**2**	**3+**	
Tortuous (%)	154 (51)	51 (17)	37 (12)	58 (19)	300	136 (45)	66 (22)	37 (12)	61 (20)	300
Crossed (%)	31 (10)	27 (9)	25 (8)	217 (72)	300	25 (8)	25 (8)	18 (6)	232 (77)	300
Ramified (%)	246 (82)	41 (14)	10 (3)	3 (1)	300	241 (80)	40 (13)	17 (6)	2 (1)	300
Bushy (%)	286 (95)	7 (2)	2 (1)	5 (2)	300	286 (95)	8 (3)	2 (1)	4 (1)	300
Bizarre (%)	277 (92)	10 (3)	5 (2)	8 (3)	300	278 (93)	8 (3)	6 (2)	8 (3)	300
Widened cap. (%)	182 (61)	33 (11)	28 (9)	56 (19)	299	180 (60)	31 (10)	27 (9)	61 (20)	299
Giant cap. (%)	244 (82)	17 (6)	23 (8)	15 (5)	299	245 (82)	16 (5)	19 (6)	19 (6)	299
	Mean in n/mm (SD)					**Mean in n/mm (SD)**				
Density	7.78 (2.50)				294	7.71 (2.39)				292
Density (90^{O})	7.06 (2.03)				293	7.09 (1.93)				292
	Median in microm (range)					**Median in microm (range)**				
Apex	21.75 (10.30 ,156.60)				292	22.35 (11.10 ,154.90)				292
Arterial	15.60 (6.70 ,115.0)				292	16.15 (5.40 ,84.40)				292
Venous	18.95 (8.50 ,135.70)				292	20.35 (7.00 ,116.30)				292
Total width	56.25 (31.30 ,264.90)				292	59.05 (27.10 ,253.00)				292

Table 2. Inter- and intra-observer reliability of qualitative and quantitative parameters

Reliability	**Inter-observer**		**Intra-observer**
Assessment	**First**	**Second**	**First**
	Weighted Kappa of ordered categories (95% CI)		
Architecture	0.72 (0.55 ,0.82)	0.56 (0.29 , 0.72)	0.75 (0.64 ,0.83)
Avascularity	0.61 (0.37 ,0.75)	0.57 (0.33 ,0.74)	0.74 (0.62 ,0.82)
Haemorrhage	0.56 (0.16 ,0.79)	0.62 (0.29 ,0.81)	0.73 (0.57 ,0.86)
Tortuous	0.39 (0.25 ,0.50)	0.33 (0.21 ,0.44)	0.68 (0.60 ,0.76)
Crossed	0.47 (0.27 ,0.62)	0.46 (0.23 ,0.64)	0.71 (0.60 ,0.80)
Ramified	0.58 (0.30 ,0.73)	0.66 (0.45 ,0.80)	0.73 (0.59 ,0.84)
Bushy	0.73 (-0.02 ,0.90)	0.50 (-0.01 ,0.82)	0.80 (0.56 ,0.94)
Bizarre	0.21 (0.00 ,0.32)	0.29 (0.06 ,0.40)	0.76 (0.55 ,0.92)
Widened capillary	0.75 (0.64 ,0.84)	0.77 (0.67 ,0.86)	0.87 (0.82 ,0.91)
Giant capillary	0.84 (0.71 ,0.92)	0.81 (0.63 ,0.93)	0.92 (0.87 ,0.96)
	Intra-class correlation coefficient of quantitative score (95% CI)		
Density	0.71 (0.69 ,0.81)	0.67 (0.66 ,0.78)	0.92 (0.92 ,0.96)
Density 90^{o} method	0.87 (0.86 ,0.92)	0.86 (0.85 ,0.91)	0.92 (0.91 ,0.97)
Apex	0.94 (0.93 ,0.97)	0.94 (0.92 ,0.97)	0.97 (0.96 ,0.99)
Arterial	0.88 (0.87 , 0.93)	0.89 (0.88 ,0.94)	0.96 (0.95 ,0.99)
Venous	0.91 (0.90 ,0.96)	0.93 (0.92 ,0.96)	0.96 (0.95 ,0.99)
Total width	0.94 (0.93 ,0.97)	0.93 (0.92 ,0.96)	0.98 (0.98 ,0.99)

Figure 1. Examples of some qualitative (bizarre, ramified, bushy, tortuous, crossed, haemorrhage) and quantitative (widened- and giant capillary) defined parameters. For details on quantitative assessment see figure 2

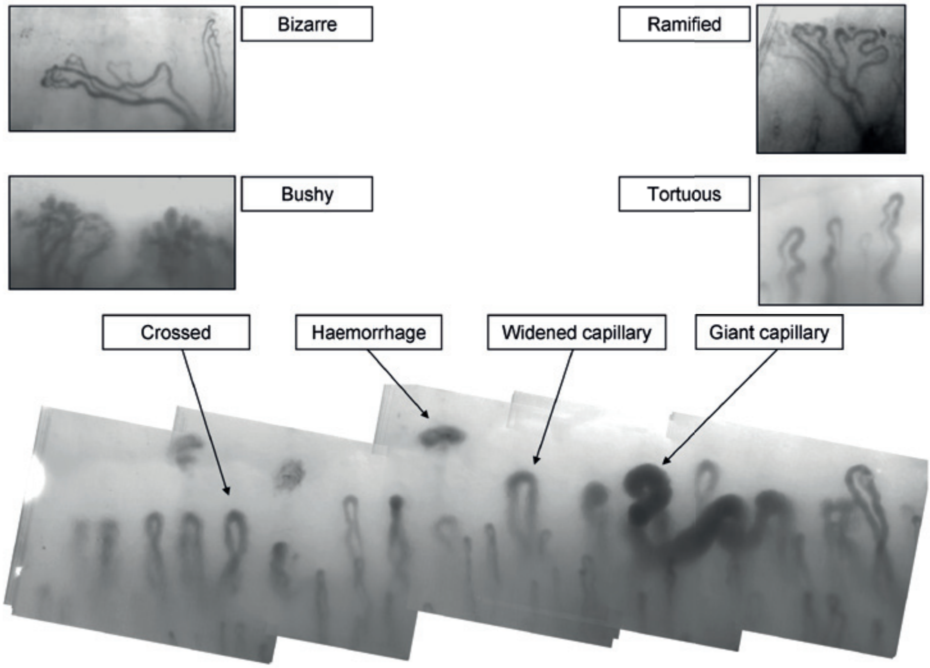
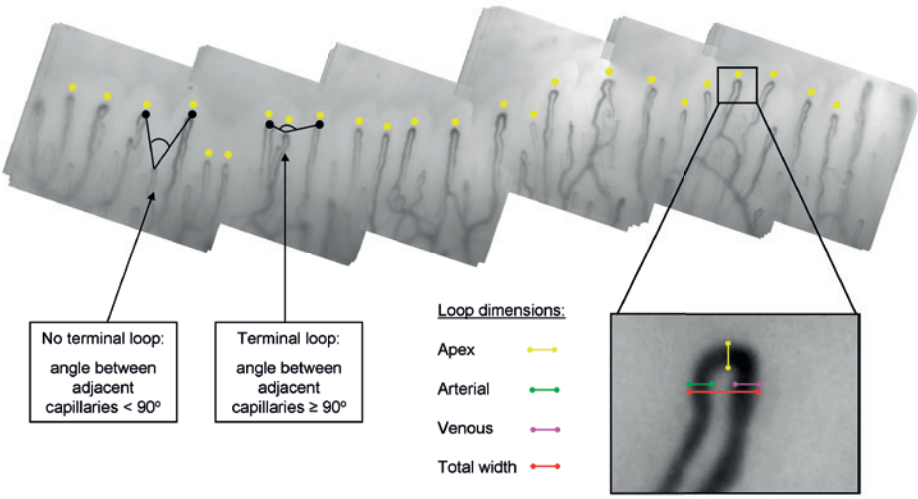


Figure 2. Measurement of capillary density by the '90° method', and measurement of capillary loop dimensions



REFERENCE LIST

1. Herrick AL, Cutolo M. Clinical implications from capillaroscopic analysis in patients with Raynaud's phenomenon and systemic sclerosis. Arthritis Rheum. 2010; 62: 2595-2604.

2. Koenig M, Joyal F, Fritzler MJ, Roussin A, Abrahamowicz M, Boire G *et al.* Autoantibodies and microvascular damage are independent predictive factors for the progression of Raynaud's phenomenon to systemic sclerosis: A twenty-year prospective study of 586 patients, with validation of proposed criteria for early systemic sclerosis. Arthritis Rheum. 2008; 58: 3902-3912.

3. Meli M, Gitzelmann G, Koppensteiner R, Amann-Vesti BR. Predictive value of nailfold capillaroscopy in patients with Raynaud's phenomenon. Clin. Rheumatol. 2006; 25: 153-158.

4. Spencer-Green G. Outcomes in primary Raynaud phenomenon: a meta-analysis of the frequency, rates, and predictors of transition to secondary diseases. Arch. Intern. Med. 1998; 158: 595-600.

5. Anderson ME, Allen PD, Moore T, Hillier V, Taylor CJ, Herrick AL. Computerized nailfold video capillaroscopy--a new tool for assessment of Raynaud's phenomenon. J. Rheumatol. 2005; 32: 841-848.6. Cutolo M, Sulli A, Pizzorni C, Accardo S. Nailfold videocapillaroscopy assessment of microvascular damage in systemic sclerosis. J. Rheumatol. 2000; 27: 155-160.

7. Kallenberg CG. Early detection of connective tissue disease in patients with Raynaud's phenomenon. Rheum. Dis. Clin. North Am. 1990; 16: 11-30.

8. Grassi W, De Angelis R. Capillaroscopy: questions and answers. Clin. Rheumatol. 2007; 26: 2009-2016.

9. Jones BF, Oral M, Morris CW, Ring EF. A proposed taxonomy for nailfold capillaries based on their morphology. IEEE Trans. Med. Imaging 2001; 20: 333-341.

10. Ingegnoli F, Gualtierotti R, Lubatti C, Zahalkova L, Meani L, Boracchi P *et al.* Feasibility of different capillaroscopic measures for identifying nailfold microvascular alterations. Semin. Arthritis Rheum. 2009; 38: 289-295.

11. Donner A, Eliasziw M. Sample size requirements for reliability studies. Stat. Med. 1987; 6: 441-448.

12. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hrobjartsson A *et al.* Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J. Clin. Epidemiol. 2011; 64: 96-106.

13. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33: 159-174.

14. Bukhari M, Hollis S, Moore T, Jayson MI, Herrick AL. Quantitation of microcirculatory abnormalities in patients with primary Raynaud's phenomenon and systemic sclerosis by video capillaroscopy. Rheumatology. (Oxford) 2000; 39: 506-512.

15. Smith V, Pizzorni C, De KF, Decuman S, Van Praet JT, Deschepper E *et al.* Reliability of the qualitative and semiquantitative nailfold videocapillaroscopy assessment in a systemic sclerosis cohort: a two-centre study. Ann. Rheum. Dis. 2010; 69: 1092-1096.

16. Cutolo M, Pizzorni C, Tuccio M, Burroni A, Craviotto C, Basso M *et al.* Nailfold videocapillaroscopic patterns and serum autoantibodies in systemic sclerosis. Rheumatology. (Oxford) 2004; 43: 719-726.

17. Matucci-Cerinic M, Allanore Y, Czirjak L, Tyndall A, Muller-Ladner U, Denton C *et al.* The challenge of early systemic sclerosis for the EULAR Scleroderma Trial and Research group (EUSTAR) community. It is time to cut the Gordian knot and develop a prevention or rescue strategy. Ann. Rheum. Dis. 2009; 68: 1377-1380.